

# 以ChatGPT技术为代表的大模型加速实现智慧运维

享受政府特殊津贴专家

全国工程勘察设计大师

国家重点研发计划重点专项专家组成员

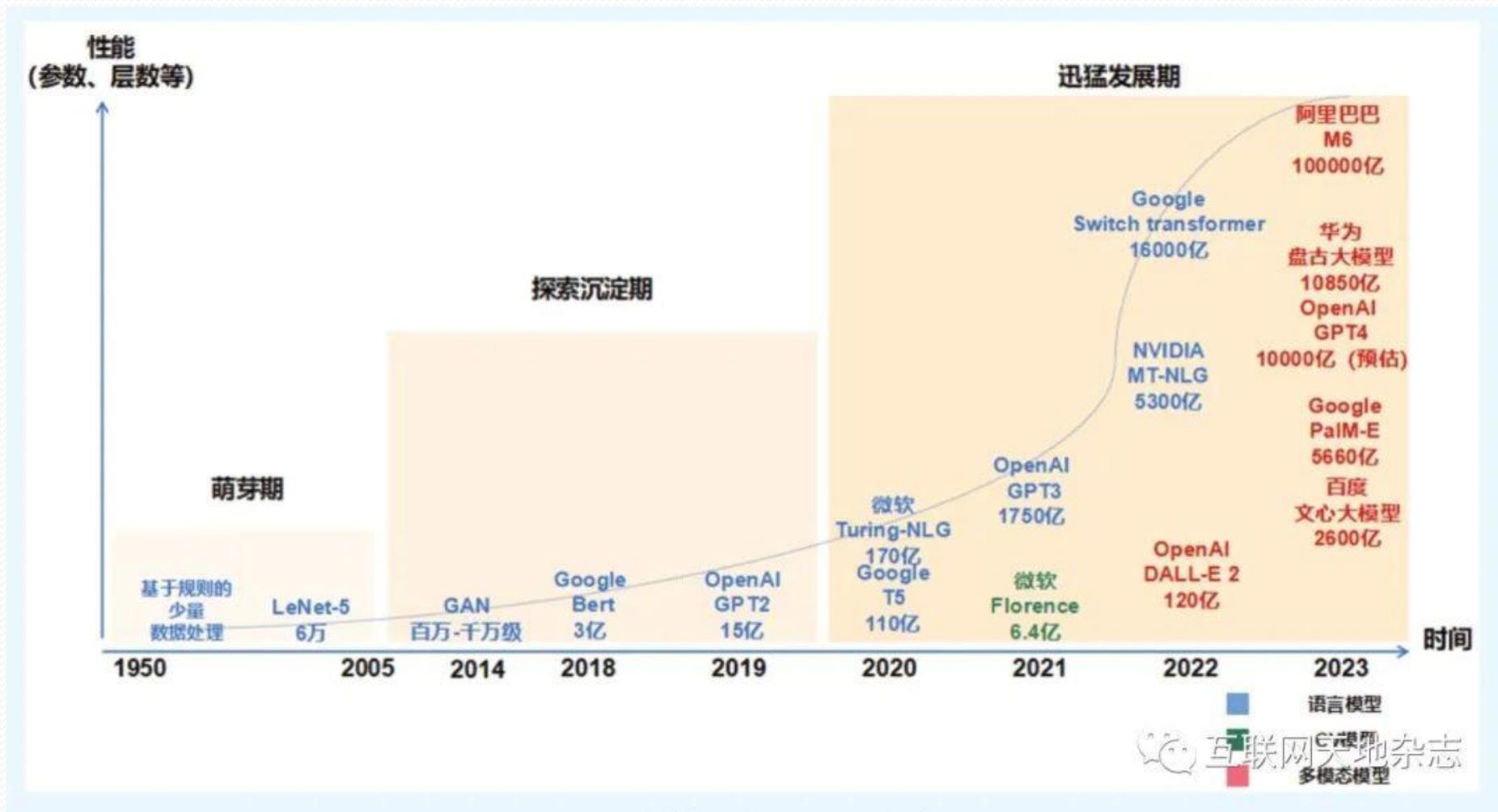
刘涛

# 以ChatGPT为代表的大模型技术加速实现智慧运维

## 目录

- 一、人工智能算法发展进程
- 二、 ChatGPT概述
  - 1、技术演进与功能
  - 2、关键技术
- 三、国内大模型发展现状
- 四、智能运维概况
  - 1、网络维护的主要内容、
  - 2、智能运维定义和现状
- 五、大模型在智能运维中应用展望
- 六、结束语

# 一、人工智能算法发展进程



## 二、ChatGPT概述

### 1、技术演进与功能概述

- ChatGPT是美国Open AI公司开发的一款基于大语言模型的聊天机器人。chat:聊天，G是generative:生成，P是pretraining:预训练，T是transformer:变换器。
- GPT家族：2018年6月份发布了GPT-1，2019年2月发布了GPT-2，2020年5月发布了GPT-3，2022年11月发布了Chat GPT，就是GPT-3.5，今年3月15日发布了GPT-4。GPT-2在GPT-1的基础上开发，GPT-3在GPT-2的基础上开发，自我升级，自我迭代。
- 模型参数GPT-1参数是1.2亿，GPT-2翻了10倍，15亿，到了GPT-3突然做了个巨大的跃迁，翻了100倍，1,750亿参数。奥秘就出在这个数量上，越过一个门槛之后，突然出现在了在门槛之下根本没有见到的现象。GPT-4的参数是1.76万亿

# ChatGTP概述

- ChatGPT已经在许多领域得到了广泛的应用，例如语言翻译、自动问答和机器写作等。
- ChatGPT具备多种强大功能，如自我纠错、生成连贯的对话响应、提供语法、拼写和修辞建议、识别并拒绝不安全的问题、提高搜索结果的相关性和准确性等。ChatGPT还具备广泛的应用领域，如问题回答、语言翻译、教育和娱乐产业等。ChatGPT代表着AIGC的未来，将在不同领域推动创新和变革
- GPT-4 在多语言理解、图文理解能力上均很强大并已融会贯通。多语言能力强大，特别是小语种能力也很出色。GPT-4在增强了安全抵御、任务完成度和图片理解能力

# ChatGTP概述

- 整体来说，GPT-4 的能力已在各种专业和学术基准上表现出了人类的水平，包括以大约前 10% 的成绩通过模拟律师资格考试，在SAT数学和阅读考试中，GPT-4比GPT-3.5高150分；它还能通过律师模拟考试，得分排在应试者前10%，GDP-4的智商145，正常人类智商是90~110之间，据说现代人智商最高的爱因斯坦160，它接近于那样一个天才的智商，这还是通识领域。
- GPT-4可以查询世界各国语言词汇、短语意思； GPT-4可以查询航班、酒店信息，帮你规划差旅； GPT-4访问各大电商数据，帮你比价甚至直接下单。另外， GPT-4接入自动化平台Zapier，与几乎所有办公软件连接，创建专属自己的智能工作流

# ChatGTP概述

- GPT-4已经可以具备了三项能力：
  - 访问最新信息，比如体育比分、股票价格、最新消息；
  - 检索知识库信息，比如公司文件、个人笔记、运行计算；
  - 代表用户执行操作，比如订机票、订餐
- 为人类提供世界知识：GPT-4 拥有庞大的世界知识，包括事实性知识和常识，相比之前的 LLM 覆盖了更多的主题和知识。不仅可以提供知识的查询和检索，而且可以进行知识的归纳和总结，甚至提供简单的知识推理和证明能力。

# ChatGTP概述

- 为人类与 LLM 之间提供了符合人类习惯的交互方式：以遵循提示（prompt）并生成补全提示词的句子的语言生成方式，通过上下文学习保持对话一致性，能理解人类意图并用自然语言回答问题、生成内容和解决问题，改变了现有人机互动方式及人类获取世界知识的方式。
- 为人类提供了跨语言及多模态交流能力：不仅可以处理多种人类语言，还可以理解图片内容，为人们提供更加便捷的交流，此外还可以将人类语言与机器语言进行相互翻译，促进了人机物三元世界的融合。
- 具备了自我学习，自我进化的能力：能够自动从海量数据和人类指令中学习到的世界知识，学习过程不需要人的介入，而且可以自行检查学习成果并不断优化迭代，能灵活应用所学知识来解决实际问题

# ChatGPT概述

## 2、关键技术

- 数据、算法和算力
- 预训练语料库的来源大致可以分为两类：  
通用数据和专业数据:如多语言数据、科学数据代码，使LLM具有特定的任务解决能力。
- 数据本身也需要花钱。训练AI必须把恶意的数据筛除，并且按照指定的格式归类。OpenAI处理这么大批量的数据数据，完全不是由程序智能归类筛选同样是由人工完成的
- GPT-4 是基于通用公开数据进行训练，缺少各行业领域的非公开的专业数据。这就造成其世界知识中缺乏深度的专业知识。因此，虽然在与人类进行沟通中，可以进行百科全书式的回答，但距离真正满足专业场景，还有相当大的差距
- GPT-4之所以独特，就是在目前世界上最强大的LLM（大语言模型）为基础的同时，引入了基于人类反馈的强化学习方法，从而提高了对话的质量，而升级到多模态预训练大模型GPT-4版本后，其对话质量变得更加强大、完美

# ChatGPT概述

- **LLM (Large Language Model, 大型语言模型)** 是指一种拥有数十亿参数的模型，通过对互联网上的大规模数据（10~100TB）进行训练而得到，可用于实现自然语言处理的多种应用。研究发现，当模型参数和用于训练的数据量达到一定规模时，LLM将出现**智能涌现**，具备一定的生成、理解和创作能力
- 大模型事实上就是算法、数据、算力上的有效结合。
- 大模型代表了新一代AI技术，带来了里程碑式的变化，解决了三个核心问题。一是，一个通用的算法就可以应对过去数十种场景，落地空间变得更大。二是，数据无需人工标注，大量数据可以得到自动训练，而且模型结构是统一的。三是，语音、图像、自然语言可以实现多模态融合，三者互相关联
- 模型能力不仅与模型大小有关，还与数据大小和总计算量有关。同时，预训练数据的质量对取得良好的性能起着关键作用，因此在扩展预训练语料库时，数据收集和清洗策略是非常重要的考虑

# ChatGPT概述

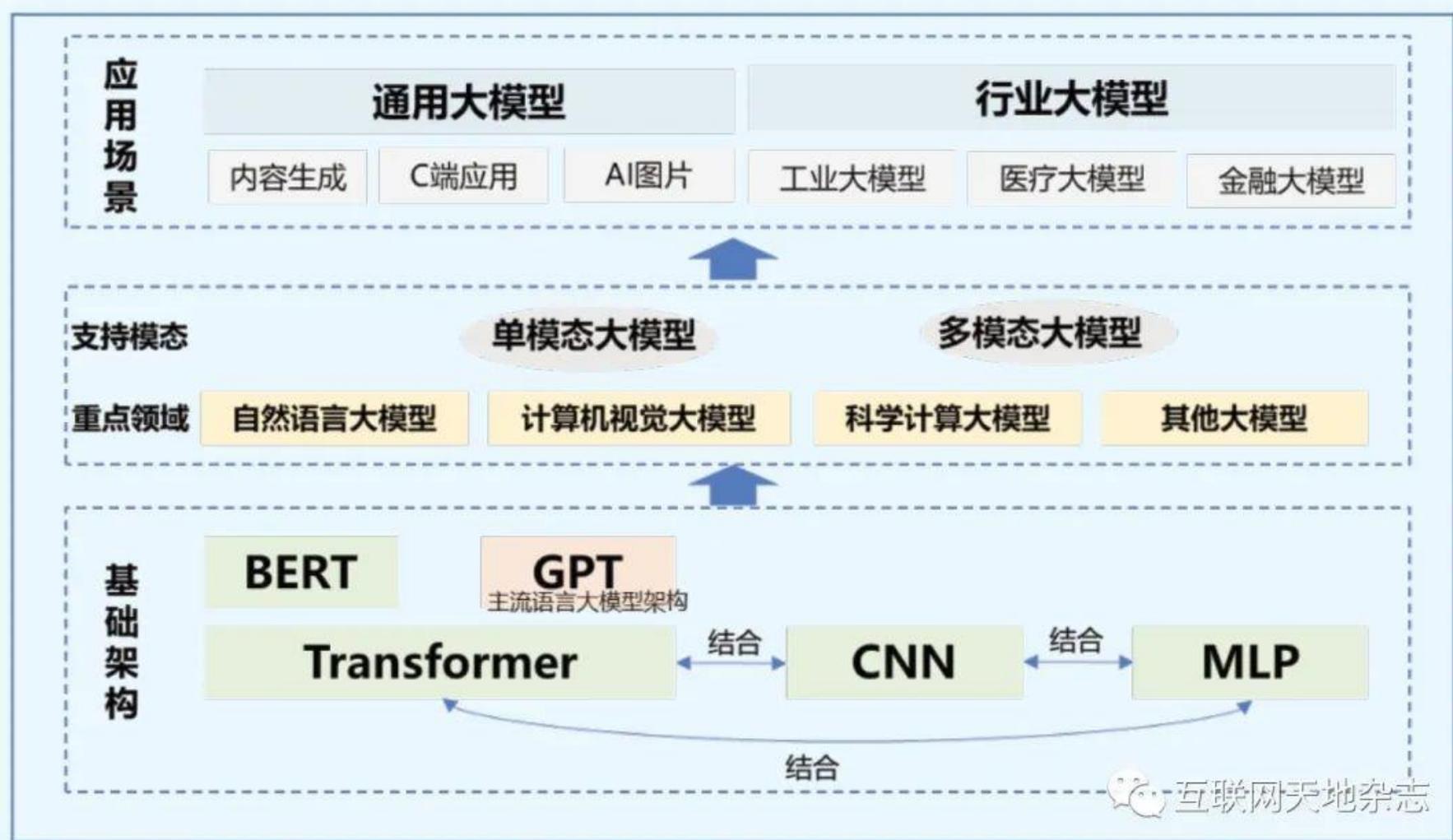
- 大模型的底层技术是谷歌的Transformer，其核心的意义在于给定前文后推理出下一个最有可能的字符来进行输出，一些巨头的基础大模型差别在于解码单向结构还是编解码双向结构。GPT-4可以支持对前文32000个字符分析再推理出下一个字，这已经是一个非常庞大的数量级，且ChatGPT只允许根据前文分析推理不允许加入后文来分析推理（decoder only）
- 大模型具有颠覆性、涌现性、通用性等特征，可以完成自然语言理解、生成、推理等多种任务；也可以处理多模态数据，如图像、视频、音频等，并通过强化学习进行自我优化
- 这种大模型算法的实现，必须有高效率的算法框架来支撑，因为参数量上去之后，需要把模型和数据分散放到多个GPU卡上，卡之间如何通信、调度，如何进行高效的反向传播，都需要大量高水平技术人员来推进和实施

# ChatGPT概述

- GPT-4的算法，是已经十分成熟的公开算法，本身没有任何秘密可言。但是这种超大规模训练，并不仅仅依靠算法本身，硬件和数据才是基础
- 即使其他团队的算法、数据、算力都准备的与OpenAI相差无几，但就是没想到以一种精巧的方式把这些元素组装起来。即使OpenAI给出了算法上的一条路径，后来者想复现GPT-4，算力、工程、数据，每一个要素都需要非常深的积累

# ChatGPT概述

- 大模型两条路径发展：第一条是通用类大模型持续拓展应用领域，打造跨行业通用化的人工智能能力平台，其应用行业正在从办公、生活、娱乐向医疗、工业、教育等加速渗透；第二条是垂直领域的专业类大模型不断深化落地，一批针对垂直领域的专业大模型提供针对特定业务场景的高质量专业化解决方案
- 多模态具身智能，是探索AGI的重要内容。GPT-4仅限于文本输入输出，其所犯的许多错误也说明了它对语义、知识和因果推理的严重缺乏
- 大模型适用于需要一定容错率的行业，大模型应用在要求100%准确的行业目前难度比较大，适用于一定的容错率而非确定性的决策工作



# ChatGPT概述

- 算力是通过对信息数据进行处理，实现目标结果输出的计算能力
- 算力主要包括三种类型，分别是基础算力、智能算力和超算算力。基础算力主要用于传统的计算应用，如计算机科学、数值计算和物理模拟等领域，智能算力是指基于GPU、FPGA、ASIC等可以加速AI计算的服务器平台提供的算法，主要用于人工智能应用，如图像识别、语音识别和自然语言处理等领域
- 智能算力和通用算力采用了不同的计算方法。以CPU为代表的通用算力，善于执行逻辑指令进行高效率的计算。GPU为代表的智能算力，善于处理大量无逻辑关系数据的并行计算

# ChatGPT概述

- 根据微软官方的数据，ChatGPT从无到有消耗的算力是这么多3640PF-days，就是说即便每秒进行千万亿次计算，也要花3640天。训练ChatGPT的庞大算力，由微软Azure的超算提供，配置至少有，285000个CPU，10000个GPU，每秒400G的网络。
- 其中这台超级计算机拥有28.5万个CPU核心，超过1万颗GPU（英伟达V100 GPU）；按此规格，如果自建IDC，以英伟达A100 GPU芯片替代V100 GPU芯片，依照性能换算，大约需要3000颗A100 GPU芯片

### 三、国内大模型发展现状

- ChatGPT官方并不支持来自中国地区的注册和访问
- 对于很多中国用户来说，因为要实现ChatGPT的功能需要大量数据，所涉及的业务也大多是数据密集型产业，然而数据出海、数据安全等问题，对于多数国内企业来说尤为敏感
- 业界公认中文互联网数据质量堪忧。从数据质量来说，整个互联网的中文数据质量，相比于英文还是有明显差距，并呼吁要想办法，做中英文不同语言之间的数据互补
- 国内AI领域真正缺乏的是技术积累，包括数据的清洗和标注、模型的结构设计，以及怎么训练、推理等，这些技术很多都需要经验和积累，需要大量的科研人才和普通技术人员一步一个脚印走出来
- 国内已经宣布布局ChatGPT类产品的科技大厂中，能够最终做出类ChatGPT产品的，最大可能还是拥有国内最成熟的大模型开发工具和产品体系的企业

## 国内大模型发展

主流厂商大模型一览			
名称	大模型产品	特征	发布时间
	文心一言	ChatGPT后 国内率先大模型场景和应用落地	2023年3月1日
	通义千问	首个超大规模语言模型	2023年4月11日
	通义听悟	聚焦音视频内容AI新品,国内首个开放公测的大模型应用产品。	2023年6月1日
	讯飞星火认知大模型	以中文为核心的新一代认知智能大模型。	2023年5月6日
	讯飞星火1.5	升级了包括开放式问答、逻辑推理和数学能力、多轮对话能力	2023年6月9日
	360智脑	千亿级参数通用语言大模型。包括生产与创作、多轮对话、代码能力、阅读理解、翻译、多模态(文本生成图像)等关键能力	2023年6月13日
	鹏城盘古	业界首个千亿级生成和理解中文NLP大模型	2021年
	HunYuan-NLP-1T	万亿级别中文NLP预训练大模型	2022年10月
	行业大模型精选商店	依托腾讯云TI平台打造,为客户提供MaaS (Model-as-a-Service)一站式服务,助力客户构建专属大模型及智能应用	2023年6月19日
	言犀产业大模型	在2021年 推出了十亿级模型K-PLUG, 在2022年 推出了百亿级模型Vega,即将推出的言犀产业大模型将达到千亿级参数,聚焦零售和金融两个优势场景	待发

## 国内发展情况

- 国内现在叫做“百模大战”，科大讯飞、昆仑万维、华为、达摩院、网易、阿里巴巴、百度、商汤、清华大学、腾讯、智源研究院，中科院等等。大模型是参数在百亿以上的，应该是只有少数大厂能做的起，它的投资强度大约是大几十亿人民币起，而且对数据的要求非常高，前期投入、算力支持、核心团队、语料支撑，同时具备这几个条件的大概中国只有那几个大厂，百度、阿里、字节、腾讯、华为等等有可能。高等院校科研院所，中科院自动化所、清华、复旦、哈工大都在做，但在算力、数据和工程化方面有短板。创业公司更适合于做垂直领域的小模型

# 国内发展情况

- 中国企业对外发布的大模型普遍是单模态的，且普遍面临GPU算力不足的挑战
- 中美在算力方面的差距，由于GPU芯片等问题，在一定程度上国内算力已被卡脖子了。即使国内头部公司，算力上的差距也是比较明显的
- 很多大企业有钱买设备，但是现在的问题是有钱买不到，现在市场上GPU卡很紧缺。目前国内云厂商拥有的GPU主要是英伟达中低性能产品（如英伟达A10）。拥有超过1万枚GPU的企业不超过5家，其中拥有1万枚英伟达A100芯片的企业最多只有1家。其中还有不少A10这样低端产品

# 四、智能运维概况

## 1、网络维护的主要内容

- 网络运维包括对网络整体表现、产品运营表现、业务使用体验、
- 对网络资源健康度进行管理、监控、分析，通过被动的监控和处理
- 通过对故障报警和性能劣化的主动感知分析以及自动化的资源调整实现网络、业务的恢复
- 通过售前、售中、售后的端到端支撑能力，提供贯穿于运维各项生产环节的自动化运维感知和决策信息的流转能力

# 网络运维基本概念

- 早期的电信网络由于架构较为简单，组网规模也不是很大，网络管控主要聚焦于设备的监管方面
- ITU-T定义了电信管理网络的五大基本域，故障（Fault）管理、配置（Configuration）管理、计费（Accounting）管理、性能（Performance）管理和安全（Security）管理简称FCAPS
- 随着SDN和NFV技术的出现，传统面向FCAPS五元组管理被控制（Control），编排（Orchestral）管理、策略（Policy）分析（Analysis）。新五元组所替代简称COMPA
- 在未来网络的管控体系中，将传统模式下的网管功能与基于SDN的控制功能融为一体，实现对辖区及辖区网元及网元连接的集中统一管控，从全网视野协调与控制所有网元的活动，统一调度网络资源、处理网络状态变化，使网络始终处于正常高效的服务状态，保证网络服务能力和质量的持续性

# 网络运维基本概念

- 网络拓扑管理
- 网络资源管理
- 业务配置管理
- 业务自动割接
- 业务生存保障
- 网络告警管理
- 网络性能管理
- 网络流量管理
- 网络自动巡检

# 网络运维面临的挑战

- 业务种类繁多，导致网络和场景更复杂，网络规模巨大，不同业务对网络性能、运维要求差异很大
- 未来网络运维将以数字化体验为中心，以数据为驱动向更自动、更智能演进
- 网络运维模式转变向“人机协同”转变
- 需要运维支撑灵活、稳定、差异化的服务
- 要求运维更敏捷、更高效、更智能、更安全
- 要求运维以服务为中心，确保提供差异化、确定性体验

# 网络运维面临的挑战

- 业务难感知
- 除了要维护物理网络外，还需要额外维护一张虚拟网络，而面向告警的运维能力是没有办法覆盖整个层面的内容
- 传统运维模式，运维人员往往扮演着事后“救火”的角色，需要事故发生后才处理。无法对未来可能发生的故障进行动态预测，掌握运维的主动权
- 故障难定位
- 管理规模大：云计算场景下运维人员的管理对象从物理设备延伸到虚拟机，网元管理规模增加了几十倍；另一方面由于实时性分析的要求，设备指标的收集粒度从分钟级提升到毫秒级，数据量增加了近千倍；更重要的是对于故障的主动感知和排障，除了收集分析网络设备指标外，还需要结合实际转发业务流进行分析，数据规模则进一步扩大。
- 业务路径多：网络为了提供高可靠和高带宽，往往被设计成负载分担方式转发流量，此时会使节点间流量经由哈希算法选择路径，转发路径的可能性随网络节点数幂级增加，管理员无法确定某业务流量究竟通过网络中什么样的路径转发，传统的定位手段时间长、严重依赖运维人员的个人经验

# 网络运维面临的挑战

- 故障恢复慢
- 网络的稳定运行关系着国家信息安全和社会稳定。为了防范灾难和风险，保障业务连续性，国内外监管部门颁布了一系列业务连续性及容灾的标准。例如对于金融行业而言，若单机构单省中断半小时，那么定位为III级事故并上报银监会。对于大部分企业，业务中断会导致企业经济利益不同程度受损。这就要求遇到故障首先要在最短时间内恢复业务。
- 安全运行
- 如果业务对技术系统的安全稳定运行方面有着高要求，但是功能涉及多个系统应用，所采用的事后处置为主的运维模式，有着异常定位困难、处理效率低等缺陷，这种被动异常响应模式不能满足异常快速定位和处理的要求
- 人力紧缺
- 由于工作内容枯燥、工作量巨大，运维岗位对人才的吸引力不高。运维需求不断上升，而人力资源依旧紧缺，这成为了技术系统发展中无法避免的矛盾

# 四、智能运维概况

## 2、智能运维的概念和现状

- 智能运维（AIOps）将人工智能应用于运维领域，基于已有的运维数据（日志、监控信息、应用信息等），通过人工智能的方式来进一步解决自动化运维无法解决的问题。
- 不依赖于人为指定规则，由人工智能算法自动地从海量运维数据中不断地学习，不断地提炼并总结规则。
- 智能运维以系统所采集的运维大数据为基础，利用人工智能算法对运维数据进行深入分析

# 智能运维的概念

- 智能运维的核心价值就在于由人工智能取代人力决策，快速给出故障处理建议，或者提前规避故障
- 智能运维是以大数据平台和算法平台为核心。
- 智能运维需要从各个监控系统中抽取数据、面向用户提供服务、并有执行智能运维产生决策模型的自动化系统
- 很多人把智能运维和自动化运维混为一谈，但其实两者天差地别。自动化运维，说简单点就是把那些传统上需要人来手工进行的操作自动化，主要是通过脚本实现。但其实还是人为判断故障原因，手动执行故障恢复脚本，或特定场景化的一系列自动化动作。这种自动化运维依赖人的策略和经验，人占主导地位，而系统只是代替人去具体执行
- 智能运维作为一个新型的技术领域，目前正在快速发展，越来越多的研究者将机器学习、自然语言处理、数据挖掘等技术应用于智能运维领域

# 智能运维的优越性

- 网络健康度评估方案对整个网络进行系统化的网络级的评估检测，帮助运维人员“看网识网”，提升运维效率和业务体验质量
- 故障根因快速定位，实现智能诊断
- 当前网络规模大、配置复杂、变化量大，很难快速地进行故障定位和排查。并且定位手段时间长，严重依赖运维人员的个人经验。网络智能运维，可以实现故障根因快速定位
- 故障自闭环，业务运行“0”中断
- 智能运维系统采用规则引擎、智能化引擎、知识图谱等技术进行大数据挖掘分析，实现对故障的快速发现和定位，并通过与控制器的联动，可实现对故障的一键式恢复或隔离。在此过程中还可根据具体故障信息给出对网络或业务的影响分析，在下发恢复或隔离预案之前，也会向用户展示对应预案下发后对网络或业务将产生哪些影响，以使用户进行决策

# 智能运维的能力

- Gartner定义的智能运维平台拥有11项能力，这些能力是对智能化运维的概念描述：
- 历史数据管理(Historical data management)
- 流数据管理(Streaming data management)
- 日志数据提取(Log data ingestion)
- 网络数据提取(Wire data ingestion)
- 算法数据提取(Metric data ingestion)
- 文本和NLP文档提取(Document text ingestion)
- 自动化模型的发现和预测(Automated pattern discovery and prediction)
- 异常检测(Anomaly detection)
- 根因分析(Root cause determination)
- 按需交付(On-premises delivery)
- 软件服务交付(Software as a service)

# 智慧运维的功能

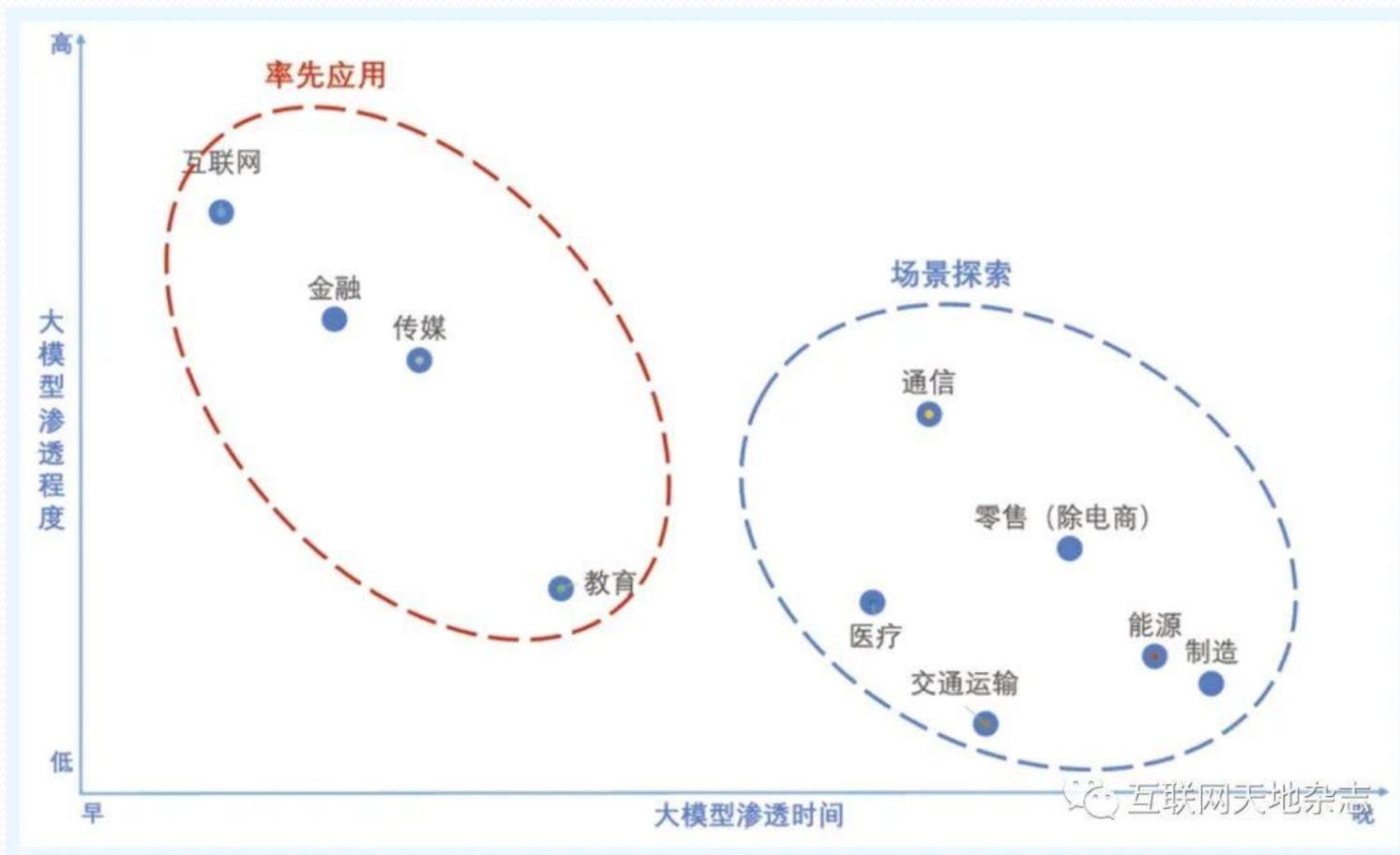
- 网络异常智能感知
- 故障智能溯源
- 健康分析
- 网络流量预测
- 生成各种运维场景

## 五、大模型在智能运维中应用展望

### 1、智慧运维应用大模型技术优势

- 近年来，深度学习成为推动智能运维发展的重要驱动力，但随着深度学习技术进入瓶颈期，产业发展面临多个挑战。
- 传统深度学习多有监督训练，对标注数据依赖性大，非常耗费人力成本；其次，模型的领域特性强，跨领域迁移能力较弱，通用性不强。此外，目前深度学习对人类感情的理解还停留在浅层次的语义层面，不具备良好的逻辑推理能力，无法真正理解用户意图。
- 大模型为代表的生成式 AI 有大算力、大数据、大模型的典型特点，首先，其基于大规模无标注数据进行预训练，再通过少量标注微调，大幅降低了对数据标注要求；其次，生成式 AI 的预训练大模型具有多模态（文字、图片、程序、视频等）、跨模态（“文生文”“文生图”“图生图”）内容生成能力，有较强的通用性和跨场景使用能力。此外，与传统 AI 相比，生成式 AI 能够更有效地捕捉用户的意图，以理解上下文，使对话更流畅自然，具有更强的逻辑和组织能力
- 大模型具有更渊博的知识和更强大的解决问题的能力，它可以以更高的精度分析解决复杂的问题

# 大模型行业渗透示意图



# 智慧运维应用大模型技术优势

- 生成故障诊断报告，提供对网络故障的快速分析，帮助网络运维人员快速识别告警的重要性并给出建议，可快速自动生成各类报告及数据可视（代码），帮助运营人员更好地理解网络数据，并生成自动化脚本，实现故障自愈。
- 借鉴大模型训练方法，可以对目前网络设备日志异常检测算法进行增强。
- 基于网络设备硬件告警、性能指标、用户投诉等多维数据，通过大模型知识引擎给出隐患预警及并识别网络存在的问题，对无线网络覆盖、容量、干扰、能耗IP路由及业务路径调整、网络云计算存储优化等常见问题，根据网络的各类数据通过知识引擎给出较为精确智能的优化方案

# 智慧运维应用大模型技术优势

- 运维人员，借助大模型，无需进行复杂的构建、部署配置，集成和发布活动变得更轻松。应用的构建部署配置从以天为单位，降低到以小时为单位，集成和部署效率提升 10 倍以上
- 大模型可以帮助自动生成测试用例，特别是在一些较为简单的场景中，可以避免手动编写重复的测试用例，提高测试的自动化程度，从而提高测试效率
- 大模型可以通过学习历史缺陷数据，识别并智能检测出潜在的缺陷，减少漏测和误测的情况
- 大模型测试还可以根据缺陷类型和严重程度进行排序，提高缺陷修复的优先级和效率
- 大模型可以基于领域知识和历史测试数据，辅助测试人员制定更加精准的测试计划，包括测试范围、测试用例设计、测试覆盖率等方面，以提高测试的效率和覆盖度，减少漏测和误测的情况

# 智慧运维应用大模型技术优势

- 通过大模型智能引擎可以通过问答方式或数据交互方式快速得到网络节能优化策略以及网络能耗vs网络性能vs用户感知的分布地图，帮助运营商有效决策及如何动态调整网络资源达到能耗&性能最优。
- 大模型可以较为完美的代替目前的知识百科、专家经验库、甚至知识图谱等系统及技术，通过问答的方式展示云网优化的专业知识和优化案例，用户可以自由就关心的问题提问，并得到较为准确的回答，同时可以较为高效的成为专业技能培训系统，短时间快速提升员工技能。
- 大模型可以有效管理各类用户需求，并及时导入专网运营平台，提升用户感知，面向来自客户域的投诉工单，基于大模型形成面向投诉的智能问答系统，识别用户投诉的语义、情感，并形成处理策略。结合RPA技术进行后续处理、关闭等自动化流转。
- 基于大模型形成面向投诉的智能问答处理系统，识别用户投诉的语义、情感，并形成处理策略，结合RPA技术进行后续处理、关闭等自动化流转，根据客户投诉内容安抚客户，对于用户的投诉，识别出网络问题后给出较为精确的定界定位的分析，并进行针对性的处理，如派单修复套餐推送等切实解决投诉的问题。

# 智慧运维应用大模型技术优势

- 行业数据的优势
- 行业算力的优势
- 行业算法模型的优势
- 行业专家的优势

# 智慧运维与智能运维

- 智慧运维与智能运维的最大区别就是智慧运维比智能运维更加智能，更节省人力
- 智慧运维除了能够比智能运维更精准地进行分析和预测外，还可以准确地生成各种场景
- 大模型的广泛应用，是实现智慧运维的重要基础和前提条件
- 智慧运维既可以基于智能运维，也可以直接实现

## 2、智慧运维中应用大模型途径

- 使用大模型建立智能运维平台，采集有关海量数据，采用大模型技术，提供智慧运维服务
- 使用大模型建立智能运维工具，专注提供模块化智慧运维服务

# 大模型在智慧运维中应用

- 高质量数据是网络的本源，网络在运行过程中会产生大规模、复杂类型的数据，除了这些数据外，还包括通信行业数据，这些数据来自于内部收集、公共数据集和第三方数据提供商
- 资源配置数据
- 客户业务数据网络，承载的客户业务信息
- 运行状态数据

# 大模型在智慧运维中应用

- 运维操作数据
- 来自于第三方系统，如光缆哑资源信息故障、工单信息、机房环境温度信息等
- 通信行业属性数据
- 客户服务对话数据。
- 行业标准和法规文件
- 通信技术数据

# 大模型在智慧运维中应用

- 对这些数据进行清洗，去除重复、冗余、不一致的数据。提取高质量的核心网络数据集
- 算法模型是挖掘网络知识，实现网络智慧的关键
- 由于从零开始训练一个大模型需要大量的数据资源和算力资源，行业大模型通常基于已有的通用大模型继续训练得到。可基于开源开放的通用大模型进行训练
- 强大算力是智慧网络的基础，由需要处理的数据量和算法决定

## 大模型在智慧运维中应用亟待解决的问题

- 海量数据的清洗和标注，特别是数据标准化
- 急需专门人才，以解决模型的结构设计，以及怎么训练、推理
- GPU智算能力不足
- 如何处理好与数据安全的关系

# 结束语

- 大模型技术将极大地促进智慧运维的发展
- 任重道远